Statistics and Probability Letters 144 (2019) 69–73

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

The scale enhanced wild bootstrap method for evaluating climate models using wavelets

Snigdhansu Chatterjee

School of Statistics, University of Minnesota, United States

ARTICLE INFO

Article history:
Available online 23 August 2018

Keywords:
Wild scale enhanced bootstrap
Climate model
Reanalysis
Wavelet
Martingale
Indian monsoon

ABSTRACT

A novel resampling-based method is presented for testing if a Physics-based model or reanalysis object successfully emulates climate signals. We establish asymptotic consistency, present simulations and a real-data application to illustrate the performance of the proposed method.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Three different kinds of data products are typically used in many studies on Earth's climate. First, observations are obtained on climate attributes of interest, like temperature, wind speed or precipitation. Second, Physics-based models of climatic processes are constructed from first principles, which are complex systems of differential equations. Such physical models are run under different initial and boundary conditions and physically meaningful choices of various tuning parameters, to generate numeric model runs of climate. The third commonly used data object in climate-related studies is *reanalysis* data, which is produced using a systematic data assimilation process using both physical climate models and observations, see Dee et al. (2016). The objective of constructing and studying Physics-based models and reanalysis objects is to understand the physical systems that govern the planet's climate, and to this end it is essential to have principled means of comparing the outputs of these models to Earth's actual *climate*. Climate is defined as long-term weather pattern, see for example NASA (2015). Note however, that the term of "long term weather pattern" is imprecise. Braverman et al. (2017) essentially consider *climate* to be the coarse level components from a wavelet decomposition of the observed weather data, and we will use that definition of climate in this paper.

We propose a hypothesis test for comparing such climate signal components as modeled by a Physics-based system or a reanalysis object to the one elicited climate data. This approach has been recently adopted in Braverman et al. (2017). In that paper, the authors first used the same wavelet decomposition of observed data on global temperature and various climate model outputs on the same variable. Then, a test of hypothesis was conducted to check if the coarse level wavelet coefficients from the observed data and model-output runs are equal, using a resampling scheme based on Autoregressive Integrated Moving Average (ARIMA) modeling of the residuals after the coarse-wavelet-based terms have been removed from the time series. Note however, Braverman et al. (2017) did not study any statistical properties of their methodology, they used a standard time series resampling scheme whose consistency was also not studied, and they used a test statistic that may not be powerful for comparing reanalysis runs to observed data. In this paper, we address these issues by presenting a full theoretical analysis, a different resampling scheme and a different hypothesis test. The methodology used in this paper

E-mail address: chatt019@umn.edu.

https://doi.org/10.1016/j.spl.2018.07.020
0167-7152/© 2018 Elsevier B.V. All rights reserved.

and in Braverman et al. (2017) may be considered complementary to each other, and the techniques used for the theoretical analysis in this paper can potentially be used in the context of Braverman et al. (2017) as well.

Additionally, our technical assumptions about the data are quite weak. This is because statisticians have very limited or no control on generation or collection of the three kinds of data products described above; such data are outputs of very complex systems and are unlikely to satisfy simplistic assumptions and are routinely verified to violate assumptions of off-the-shelf popular statistical procedures; the goal of climate systems modeling is to understand the complex dependence structure among its constituent elements, and simplistic assumptions to enforce a statistical requirement may be misleading in the study of the Physics of these systems.

In view of this, we propose a resampling scheme to implement the hypothesis test of comparing climate signals of observations and model outputs. The resampling scheme we use is a variant of the *wild bootstrap* (Mammen, 1993), where we impose the additional condition that the variance of the resampling weights tending to infinity with the sample size, thus a *wild, scale enhanced* bootstrap procedure is used. We establish the size and power properties of the test, and consistency of the proposed new resampling technique for the testing problem. The asymptotic analysis of the properties of the test statistic is not standard; for example, the speed of convergence depends both on the sample size and the number of wavelet coefficients used. The asymptotic consistency of the resampling procedure has additional complications.

Simulation experiments and a real data example are used to illustrate the methods proposed in this paper. The simulation results essentially show that our proposed test methodology maintains size (probability of Type-I error), and is in fact slightly conservative, and that it has good power properties. We implement our proposal on precipitation data from the Indian subcontinent, where modeling reliable precipitation is a recognized challenging problem, references and details are provided later. We note that there has been some attempts to compare climate model runs to observed data, sometimes visually and in an *ad hoc* way, sometimes in terms of trends or other broad features, typically under restrictive statistical assumptions. As a small sample of the related literature we refer to Murphy et al. (2004), Hemer et al. (2013) and Braverman et al. (2017).

The layout of the paper is as follows: in Section 2 we describe in detail the statistical model and the methodology we use to define climate signals, and then conduct the hypothesis test to compare signals from the observed data and Physics-based models. We discuss technical assumptions and notations also in this section. In Section 3 we present the main theoretical result on the consistency of the wild, scale enhanced bootstrap for estimating the null distribution of the test statistic, along with its proof. In Section 4 we present a simulation exercise to illustrate the power properties of our proposed test in a variety of situations. In Section 5 we present a case study on precipitation in the Indian subcontinent. Throughout this paper we use the conventions that all vectors are column vectors, a^T denotes the transpose of a (vector or matrix) a, the notation $\langle a, b\rangle$ denotes the Euclidean inner product of two vectors a and b, and the Euclidean norm is denoted by $\|\cdot\|$.

2. The model

We denote the observed data by $\widetilde{\mathbf{Y}}_0 = (\tilde{Y}_{01}, \ldots, \tilde{Y}_{0n})^T \in \mathbb{R}^n$, and the ℓ^{th} instance of a reanalysis data or climate model run by $\widetilde{\mathbf{Y}}_\ell = (\tilde{Y}_{\ell 1}, \ldots, \tilde{Y}_{\ell n})^T \in \mathbb{R}^n$, $\ell = 1, \ldots, L$. We assume that the data are discrete observations from the data curves $\tilde{Y}_\ell(t)$ for $t \in [0, 1]$, with the identification $\tilde{Y}_{\ell i} \equiv \tilde{Y}_\ell\big(i/(n+1)\big)$, $\ell = 0, 1, \ldots, L$. We assume that the climate signature components inherent in $\tilde{Y}_\ell(\cdot)$ are elements of the up to measure zero equivalence class of functions forming the Hilbert space $\mathcal{L}_2([0,1]) = \{f : [0,1] \to \mathbb{R}; \int_{x=0}^1 f(x)^2 dx < \infty\}$. Further, suppose $\{V(t), W_{jk}(t), j = 0, 1, \ldots, k = 0, 1, \ldots, 2^j - 1\}$ be a family of wavelet basis functions of $\mathcal{L}_2([0,1])$. See Antoniadis (1997, 2007) for reviews of wavelet methods in Statistics, with many additional references. We first isolate the trend, seasonality, and known dependence on covariates. For example, in the simulations and the application on precipitation data in India later in this paper, we first fit a linear trend, then a seasonality component, and extract the residuals which are then modeled using the wavelet basis system described above. Thus, our working dataset $\{Y_\ell(t), \ell = 0, \ldots, L, t = i/(n+1), i = 1, \ldots, n\}$ may be derived from the original datasets $\{\tilde{Y}_\ell(t)\}$, and we write $Y_\ell(t) = \mu_\ell(t) + \epsilon_\ell(t)$, where we assume that

$$\mu_\ell(t) = \sum_{j=0}^{J_{n0}-1} \sum_{k=0}^{2^j-1} \gamma_{\ell jk} W_{jk}(t). \tag{1}$$

Naturally, it is possible to include the trend and other components in the model with a more complex formulation of (1), we do not do that for simpler notations and clearer presentation of the main concepts. Note that for simplicity we have tacitly assumed that the scaling function $V(\cdot)$ has no contribution in $\mu_\ell(\cdot)$. This is often the case when $Y_\ell(\cdot)$ are residuals after fitting trend or other functionals to the original data, and in any case, including it in $\mu_\ell(\cdot)$ or in the trend of the original data has little effect on the rest of this paper.

The value of J_{n0} in (1) is specified using knowledge of the Physics driving the climate feature under study. The exact mathematical requirement on J_{n0} is given later in this paper, and we allow it to depend on sample size n, thus allowing considerable flexibility. The function $\mu_0(\cdot)$, being based on the coarse scales $j = 0, \ldots, J_{n0} - 1$, contain the signatures of the irregularly periodic long-term climate signal components like *El Nino*-Southern oscillation (ENSO), decadal and multi-decadal oscillations.

Define the $p_n = 2^{J_{n0}} - 1$ dimensional vectors $\mathbf{x}_i$, whose elements are $W_{jk}(i/(n+1))$ for $j = 0, \ldots, J_{n0} - 1, k = 0, \ldots, 2^j - 1$. We collect these vectors in a $n \times p_n$ matrix X, whose ith row is $\mathbf{x}_i$, $i = 1, \ldots, n$. Further define $\boldsymbol{\gamma}_\ell$ as the p_n-dimensional vector consisting of the $\gamma_{\ell jk}$'s, and $\mathbf{Y}_\ell = (Y_{1\ell}, \ldots, Y_{i\ell}, \ldots, Y_{n\ell})^T \in \mathbb{R}^n$. Also define $\epsilon_{\ell i} \equiv \epsilon_\ell\big(i/(n+1)\big)$, and collect these in the vector

$\epsilon_\ell = (\epsilon_{1\ell}, \ldots, \epsilon_{n\ell})^T$ for $\ell = 0, 1, \ldots, L$. In terms of these, we may write the data model as $\mathbf{Y}_\ell = X\boldsymbol{\gamma}_\ell + \epsilon_\ell$. Projecting $\boldsymbol{\gamma}_\ell$ on the space spanned by $\boldsymbol{\gamma}_0$ results in the relation

$$\boldsymbol{\gamma}_\ell = \beta\boldsymbol{\gamma}_0 + \delta, \text{ where } \beta = \left[\|\boldsymbol{\gamma}_0\|^2\right]^{-1}\langle\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1\rangle, \text{ and } \langle\boldsymbol{\gamma}_0, \delta\rangle = 0. \tag{2}$$

Thus for a known value β_*, the hypothesis of interest in this paper is

$$H_0 : \beta = \beta_*, \quad H_1 : \beta \neq \beta_*.$$

The case $\beta_* = 1$ is the one of interest, however, we develop the methodology below for any fixed β_*.

We obtain the estimates of the slope parameters using ordinary least squares estimation technique on $\mathbf{Y}_\ell$, thus we have $\hat{\boldsymbol{\gamma}}_\ell = \left(X^TX\right)^{-1}X^T\mathbf{Y}_\ell = \boldsymbol{\gamma}_\ell + \left(X^TX\right)^{-1}X^T\epsilon_\ell$. Let $\mathcal{A}_n$ be the set on which $\|\hat{\boldsymbol{\gamma}}_0\|^2 > \epsilon_n = 0.5c_0p_n$, thus $\mathcal{A}_n = \{\|\hat{\boldsymbol{\gamma}}_0\|^2 > \epsilon_n\}$. We define the estimator of β as follows: $\hat{\beta} = \langle\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\gamma}}_1\rangle/\|\hat{\boldsymbol{\gamma}}_0\|^2\mathcal{I}_{\{\mathcal{A}_n\}}$, where $\mathcal{I}_A$ is the indicator of the set A, that is, it is 1 in A and 0 otherwise.

We make the following technical assumptions about our model:

$$|\boldsymbol{\gamma}_{0j}| \in \left(c_0,\ C_0\right), \text{ for two constants } 0 < c_0 < C_0 < \infty. \tag{3}$$

$$\|\boldsymbol{\gamma}_\ell\|^2 = O(p_n),\ \ell = 1, \ldots, L. \tag{4}$$

$$\text{The eigenvalues of } n^{-1}\left(X^TX\right) \in \left(c,\ C\right). \tag{5}$$

$$\text{tr}\mathbb{E}\left(X^T\epsilon_0\epsilon_0^TX\right) = O(np_n). \tag{6}$$

$$n^{-1}p_n \to 0, \text{ and } p_n \to \infty \text{ as } n \to \infty. \tag{7}$$

Note that since $p_n = 2^{J_{n0}} - 1$, the condition (7) imposes conditions on J_{n0}. These conditions suggest that $J_{n0} \to \infty$ as $n \to \infty$, but at a rate slower than $\log_2(n)$.

Let $\boldsymbol{\epsilon}_i = (\epsilon_{i0},\ \epsilon_{i1})^T \in \mathbb{R}^2$. Let $\mathcal{F}_i$ be the σ-algebra generated by $\{\boldsymbol{\epsilon}_j; j \leq i\}$, and $\mathcal{F}_0$ defined as $\{\emptyset, \Omega\}$, the trivial σ-algebra. We assume that $\mathbb{E}[\boldsymbol{\epsilon}_i|\mathcal{F}_{i-1}] = 0$, $\mathbb{V}[\boldsymbol{\epsilon}_i|\mathcal{F}_{i-1}] = \Sigma_i$. Let $\tilde{\boldsymbol{\gamma}}$ be the $p_n \times 2$ dimensional matrix with the column vectors given by $\boldsymbol{\gamma}_1 - 2\beta\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_0$. Define the $n \times 2$ dimensional matrix $A = X\left(X^TX\right)^{-1}\tilde{\boldsymbol{\gamma}}$, and let $A_i = \mathbf{x}_i^T\left(X^TX\right)^{-1}\tilde{\boldsymbol{\gamma}}$, $i = 1, \ldots, n$ be the rows of A. Define $Z_i = n^{1/2}p_n^{1/2}\|\boldsymbol{\gamma}_0\|^{-2}A_i^T\boldsymbol{\epsilon}_i$, $i = 1, \ldots, n$. We assume that $\{(Z_i, \mathcal{F}_i)\}$ form a martingale difference sequence. That is $\mathbb{E}[Z_i|\mathcal{F}_{i-1}] = 0$, and consequently if we define $S_n = \sum_{i=1}^n Z_i$, then $\{(S_n, \mathcal{F}_n), n \geq 1\}$ is a martingale. The conditional variance $\mathbb{V}[Z_i|\mathcal{F}_{i-1}]$ is given by

$$\sigma_{ni}^2 = np_n\|\boldsymbol{\gamma}_0\|^{-4}\mathbf{x}_i^T\left(X^TX\right)^{-1}\tilde{\boldsymbol{\gamma}}\Sigma_i\tilde{\boldsymbol{\gamma}}^T\left(X^TX\right)^{-1}\mathbf{x}_i,$$

and based on these we define $V_n = \mathbb{E}\sum_{i=1}^n\sigma_{ni}^2$. We assume that $\{Z_i\}$ obeys the Lindeberg condition, and that $\sum_{i=1}^n\mathbb{E}\left|\mathbb{E}\left(Z_i^2|\mathcal{F}_{i-1}\right) - \sigma_{ni}^2\right| = o(V_n)$. We assume that

$$V_\infty = \lim_{n\to\infty}\mathbb{E}\sum_{i=1}^n\sigma_{ni}^2 = \lim_{n\to\infty}np_n\|\boldsymbol{\gamma}_0\|^{-4}\sum_{i=1}^n\mathbf{x}_i^T\left(X^TX\right)^{-1}\tilde{\boldsymbol{\gamma}}\Sigma_i\tilde{\boldsymbol{\gamma}}^T\left(X^TX\right)^{-1}\mathbf{x}_i$$

exists and is a finite quantity. Note that the above conditions ensure that a Central Limit Theorem holds for $\{Z_i\}$, see for example Chow and Teicher (1997), page 336. We note that the underlying assumptions of standard regression and time series models often used for climate data analysis imply the above technical conditions.

3. Resampling algorithm for hypothesis testing

Recall that $p_n = 2^{J_{n0}} - 1$. To simplify notations in the rest of this paper we will occasionally drop the suffix of p_n, and refer to it as p. Define the centered and scaled random variable $T_n = n^{1/2}p_n^{1/2}\left(\hat{\beta} - \beta\right)$. The aim of this section is to propose a resampling scheme, and to show its consistency, in approximating the distribution of T_n under the null hypothesis. In this section we use $\ell = 1$ as the template climate model run or reanalysis data for discussion.

Define the residuals $\mathbf{r}_\ell = \mathbf{Y}_\ell - X\hat{\boldsymbol{\gamma}}_\ell$, $\ell = 0,1$. Also let U_b be a diagonal matrix of independent, mean zero, unit variance random variables that have finite fourth moment. Let τ_n be a sequence satisfying $\tau_n \to \infty$ and $n^{-1}p^{-1}\tau_n^2 \to 0$ as $n \to \infty$. We define $\mathbf{Y}_{\ell b} = X\hat{\boldsymbol{\gamma}}_\ell + \tau_nU_b\mathbf{r}_\ell$. This yields the resample-based estimators $\hat{\boldsymbol{\gamma}}_{\ell b} = \hat{\boldsymbol{\gamma}}_\ell + \tau_n\left(X^TX\right)^{-1}X^TU_b\mathbf{r}_\ell$. Define the set $\mathcal{A}_{nb} = \{\|\hat{\boldsymbol{\gamma}}_{0b}\|^2 > \epsilon_n\}$. We then obtain the resampled estimator $\hat{\beta}_b = \left[\|\hat{\boldsymbol{\gamma}}_{0b}\|\right]^{-2}\langle\hat{\boldsymbol{\gamma}}_{0b}, \hat{\boldsymbol{\gamma}}_{1b}\rangle\mathcal{I}_{\{\mathcal{A}_{nb}\}}$. Define $T_{nb} = n^{1/2}p^{1/2}\tau_n^{-1}\left(\hat{\beta}_b - \beta_0\right)$ as the resampling version of $T_n = n^{1/2}p^{1/2}\left(\hat{\beta} - \beta\right)$ under the null hypothesis.

Theorem 3.1. *For almost all data sequences, the distribution of T_{nb} conditional on the data and the distribution of T_n converge to the same limit. Thus, the resampling distribution of T_{nb} may be used in testing the null hypothesis $\beta = \beta_0$ against the alternative that the null is false.*

In order to establish this theorem, we first prove a central limit theorem for T_n. We can then follow similar arguments to show that under the null hypothesis T_{nb} has the same asymptotic limit almost surely, which establishes Theorem 3.1.

Table 1
Table of rejection probabilities for the test of size 0.05 corresponding to signal to noise ratio approximately 1.0.

n	J_0	β		
		1	1.2	1.5
128	1	0.04	0.01	0.31
	3	0.03	0.06	0.38
	5	0.04	0.11	0.56
256	1	0.01	0.02	0.34
	3	0.01	0.03	0.46
	5	0.01	0.23	0.57
512	1	0.00	0.42	0.68
	3	0.04	0.53	0.73
	5	0.02	0.54	0.86
1024	1	0.00	0.53	0.65
	3	0.01	0.67	0.91
	5	0.03	0.73	1.0

Lemma 3.1. *Under the conditions stated above, the distribution of* $T_n = n^{1/2}p^{1/2}\big(\hat{\beta} - \beta_0\big)$*weakly converges to* $N(0, V_\infty)$.

The proofs of Lemma 3.1 and Theorem 3.1 are included in the supplementary material.

4. Simulation experiments

We conducted several simulation studies to evaluate the performance of our proposed hypothesis testing method, here we report a small selection of the results. We generate data from the two processes $Y_{\ell t} = \gamma_{\ell 0} + \gamma_{\ell 1} + \mu_{1t} + \epsilon_{\ell t}$ for $t = 1, \ldots, n$, $\ell = 0,1$. In our simulations, the series Y_{0t} acts as the "observed" series and the second one Y_{1t} acts as the "model" output. We consider ϵ_{0t} and ϵ_{1t} to be all independent and identically distributed as $N(0, V)$.

The signal components of both the processes Y_{0t} and Y_{1t} satisfy the framework we adopt in this paper, namely $\mu_{\ell t} = \sum_{j=0}^{J_0-1}\sum_{k=0}^{2^j-1}\gamma_{\ell jk}W_{0jk}(t)$, $\ell = 0,1$. The signal μ_{0t} and the noise parameters for Y_{0t} is based on a model fitted to the observed global temperature record, to relate to the context of this paper. For the series μ_{1t}, we use the relation $W_{1jk} = \beta_0 + \beta W_{0jk}$. For all the simulations reported below, we used $\alpha = 0$, $\gamma_{\ell 0} = 0$ and $\gamma_{\ell 1} = 0$. We use a grid of β values in the interval $[0, 2]$. However, all these parameters are considered to be unknown in the actual computation. The error structure in either series is assumed to be a heteroscedastic martingale difference series as in Section 2.

Our resample size (bootstrap Monte Carlo size) is $B = 200$. Each of the above scenario is independently replicated for $R = 300$ times. Our simulation code is built using programs similar to our R package `WiSEBoot` (Heyman and Chatterjee, 0000).

The null hypothesis is $H_0 : \beta = 1$ the alternative and $H_1 : \beta \neq 1$. First, we study the size and power properties our hypothesis testing procedure for sample sizes $n \in \{128, 256, 512, 1024\}$, $J_0 \in \{1, 3, 5\}$ and noise variances selected to retain the *signal to noise* ratio $SNR \in \{0.5, 1.0, 3.0\}$. The *signal to noise* ratio is defined as $SNR = \big(T + 1 - 2^{J_0}\big)\big(2^{J_0} - 1\big)^{-1}\big\{\gamma_{11}^2 + \sum_{j=0}^{J_0-1}\sum_{k=0}^{2^j-1}\gamma_{1jk}^2\big\}\sum_{t=1}^{n}\mathbb{V}(\epsilon_{1t})$. We report the results in Table 1 for $SNR = 1$, other cases are reported in the supplementary document. Our test maintains size (probability of Type-I error), and is in fact on the conservative side. The power of the test increases very noticeably with increased signal-to-noise ratio, with scale J_0, and when the sample size n increases. The power of our test is poor when the sample size is low, the strength of the signal is low compared to the noise variance, or when the true β is close to its value under the null. These are all reasonable and expected properties, and helps establish boundaries of how and when we may expect credible results from the proposed method. Further simulation studies, including a closer look at the power function, is included in the supplementary material.

5. Indian precipitation data modeling

We use the proposed wavelet decomposition followed by hypothesis testing approach to evaluate how a few climate models and reanalysis objects perform in replicating precipitation in the Indian sub-continent. We focus on the region 19-29N, 72-89E which broadly encompasses the northern plains of the Indian subcontinent. As the observed data, we use the monthly precipitation records in millimeters per day units from CRU TS3.23 Precipitation data averaged over 5×5 grid boxes, U. of East Anglia Climatic Research Unit et al. (2015). We use the monthly CMIP5 historical runs from the climate models ACCESS1_0, three ensemble members of ACCESS1_3, bcc_csm1, bcc_csm1m, BNU_ESM, four ensemble members from CANESM2. There is no scientific reason behind selecting these models; they are alphabetically the first few alternatives available. Further technical details about the climate models is available from Taylor et al. (2012). Additionally, we use the monthly reanalysis fields from NCEP/NCAR for precipitation at the surface where the data is available from 1948 (Kalnay et al., 1996), and from the NASA-MERRA (NASA, 2016) for precipitation at the surface where the data is available from 1980.

Table 2
Table of p-values, expressed as percentage, corresponding to tests of equality in climate signals from observed data and different Physics-based models. We use $J_0 = 3$.

Start Year =	1901	1948	1980
Length =	1024	512	256
ACCESS1_0_0	25.50	42.00	9.10
ACCESS1_3_0	6.95	14.30	9.00
ACCESS1_3_1	18.15	59.30	0.70
ACCESS1_3_2	12.90	13.15	37.80
bcc_csm1_1_0	1.25	12.25	10.90
bcc_csm1_1_1	5.85	14.30	41.00
bcc_csm1_1_2	10.75	37.55	12.00
bcc_csm1_1m_0	33.05	16.85	64.70
bcc_csm1_1m_1	11.40	1.35	13.85
bcc_csm1_1m_2	11.00	29.30	11.90
BNUESM_0	7.30	15.90	23.40
CANESM2_0	7.30	22.45	5.90
CANESM2_1	10.50	8.90	46.35
CANESM2_2	7.05	5.55	24.75
CANESM2_3	2.35	18.90	39.60
CANESM2_4	40.10	33.60	4.80
NCEP		77.45	17.65
MERRA			53.85

We note that the datasets may use slightly different grids and coverage region, since the gridding process for the models are unrelated to each other.

Since the datasets are of different lengths, we conduct three separate analysis, with sample sizes of $n = 1024$ starting from 1901, $n = 512$ starting from 1948, and $n = 256$ starting from 1980. We consider $J_0 = 3$, which relates to decadal time scales. We use a resample Monte Carlo size of $B = 2000$ in this section. The p-values from the hypothesis tests conducted with these models are reported in Table 2. The models bcc_csm1_1m_0, bcc_csm1_1m_2, ACCESS1_0_0 perform better than the other models, and reanalysis runs tend to perform quite well.

Acknowledgments

This research is partially supported by the National Science Foundation (NSF) under grants # DMS-1622483, # DMS-1737918, and by the National Aeronautics and Space Administration (NASA). We would like to thank the anonymous reviewers, and additionally Dr. Amy Braverman, Dr. Noel Cressie and Dr. Megan Heyman for several extremely helpful comments.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2018.07.020.

References

Antoniadis, A., 1997. Wavelets in statistics: A review. J. Italian Stat. Soc. 6 (2), 97–130.
Antoniadis, A., 2007. Wavelet methods in statistics: Some recent developments and their applications. Stat. Surv. 1, 16–55.
Braverman, A., Chatterjee, S., Heyman, M., Cressie, N., 2017. Probabilistic evaluation of competing climate models. Adv. Stat. Climatol., Meteorol. Oceanogr. 3, 93–105.
Chow, Y.S., Teicher, H., 1997. Probability Theory: Independence, Interchangeability, Martingales, third ed. Springer-Verlag, New York.
Dee, D., Fasullo, J., Shea, D., Walsh, J., N. C. for Atmospheric Research Staff (Eds.), 2016. The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables, (Accessed December 2, 2016). URL https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.
Hemer, M.A., Fan, Y., Mori, N., Semedo, A., Wang, X.L., 2013. Projected changes in wave climate from a multi-model ensemble. Nature Clim. Change 3 (5), 471–476.
Heyman, M., Chatterjee, S., Wiseboot: A r package, CRAN.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al., 1996. The ncep/ncar 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–471.
Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Ann. Statist. 21 (1), 255–285.
Murphy, J.M., Sexton, D.M., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430 (7001), 768–772.
NASA, 2015.NASA - What's the Difference Between Weather and Climate? URL http://www.nasa.gov/mission_pages/noaa-n/climate/climate_weather.html.
NASA, 2016.Modern-Era Retrospective analysis for Research and Applications, Version 2. URL https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 (4), 485–498.
U. of East Anglia Climatic Research Unit, , Harris, I.C., Jones, P., 2015. CRU TS323: Climatic Research Unit (CRU) Time-Series (TS) Version 323 of High Resolution Gridded Data of Month-by-month Variation in Climate (Jan. 1901–Dec. 2014). Centre for Environmental Data Analysis.